Monitoring of Network Topology Dynamics

Dr. Vladimir Gudkov / Dr. Joseph E. Johnson
Department of Physics and Astronomy
University of South Carolina,
Columbia, SC 29208
USA
Email: gudkov@sc.edu / jjohnson@sc.edu

Mr. Rajesh Madamanchi
Department of Computer Science and Engineering
University of South Carolina,
Columbia, SC 29208
USA
Email: madamanc@engr.sc.edu

Mr. James L. Sidoran
Air Force Research Laboratory
Defensive Information Warfare
Rome, NY 13441-4505
USA
Email: James.Sidoran@rl.af.mil

ABSTRACT:

We present software for deriving innovative metrics describing dominant parts of the internal structure of large networks. The algorithm is sufficiently fast for the network metrics to support real time monitoring of network dynamics. The network connections (connectivity matrix) are mathematically constructed by capturing the appropriate header parameters of selected internet/network traffic. Our metrics are in part derived from a network cluster decomposition that is based upon a physical model analogy for the network that is very rapidly evolved revealing cluster structures. Certain of the metrics consist in part of Renyi (generalized Shannon) entropy measures on the resulting network clusters and subclusters. This evolution depends upon the connectivity matrix and revels many qualitative features of the network.

1.0 INTRODUCTION

Network systems have very complex structures and temporal behavior because of their intensive nonlinear and convoluted information dynamics. Therefore, recent advances in the mathematical physics of complex systems, as well as computational biology, for the modeling and simulation of complex systems could provide the framework and scale level needed for the development of a quantitative foundation for cyberspace systems.

*Paper presented at the RTO IST Symposium on "Adaptive Defence in Unclassified Networks",
held in Toulouse, France, 19 - 20 April 2004, and published in RTO-MP-IST-041.*

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE
01 NOV 2004

2. REPORT TYPE
N/A

3. DATES COVERED
-

4. TITLE AND SUBTITLE
Monitoring of Network Topology Dynamics

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Computer Science and Engineering University of South Carolina, Columbia, SC 29208 USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001845, Adaptive Defence in Unclassified Networks (La defense adaptative pour les reseaux non classifies)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**44**	

Existing approaches for the study of network information traffic usually include the study of the dependence of network stability in terms of network complexity and topology (see, for example [1,2] and references therein); signature-based analysis techniques; and statistical analysis and modeling of network traffic (see, for example [3-6]). Recently, methods have been proposed to study both spatial traffic flows [7], and correlation functions of irregular sequences of numbers occurring in the operation of computer networks [8].

In this paper we describe the developed network monitoring technique based on new approaches [9-12]: one for reconstruction of network topology, using a physics analogy of the motion of particles in a liquid medium in a multi-dimensional space, and another for rapid monitoring of topology dynamics, using generalized entropies.

2.0 ALGORITHM DESCRIPTION

For network cluster decomposition we use an algorithm based on an analogue physical model which is dynamically evolved. The detailed algorithm description is given in papers [9-11]. Here we recall the main features of the algorithm. To describe the connectivity of a network with N nodes we use the connectivity matrix C $(n \times n)$, refereed to in graph theory as the adjacency matrix. We allow matrix elements C_{ij} to be only 0 or 1 - for disconnected and connected nodes i and j, correspondingly. It should be noted that if two matrices differ only by the labeling of the vertices, effecting the permutation of rows (columns), they represent the same network.

The number of C matrices representing the same network is equal to $2^{\frac{n(n-1)}{2}}$ and is very large already for moderate size of network. We are looking for the unique matrix C which has block-diagonal structure representing the active dynamically connected groups of node on the given network. To solve this problem avoiding large number ($n!$) of operations in combinatorial approach, we use a completely symmetric and unbiased initial configuration which does not depend on the numbering: place the n nodes of the network at the n vertices of a symmetric simplex inscribed inside the unit sphere in $n-1$ dimensions. All vertices are equidistant from the origin and from each other.

The distance between any pair of vertices of the simplex i.e. between any pair of the nodes is, therefore:

$$|\vec{r}_i - \vec{r}_j| = \sqrt{\frac{2n}{n-1}} \qquad \text{for all } i \neq j \quad i, j = 1...n, \tag{1}$$

We next consider the nodes as massive point-like particles and endow our system with some dynamics introducing an attractive force between points corresponding to nodes which are connected in the initial network of interest.

Thus we postulate linear forces

$$\vec{F}_{ij}(\vec{r}_i, \vec{r}_j) = g(C_{ij}) \frac{(\vec{r}_i - \vec{r}_j)}{|\vec{r}_i - \vec{r}_j|}, \tag{2}$$

where the $g(C_{ij})$ is the intensity of interactions as a function of the value of matrix element C_{ij}. To be the force attracting the point mass i to the point mass j, in the direction of $\vec{r}_i - \vec{r}_j$. To retain the initial symmetry and avoid any biasing we take the same force law for all pairs. Then the only way information about the specific graph of interest is communicated to our dynamical n body system is via the overall strengths $g(C_{ij})$ of the forces. In the presented program the default value of $g(C_{ij})$ is equal to zero if $C_{ij} = 0$.

Next let our point move according to first order "Aristotelian Dynamics"':

$$\mu_i \frac{d\vec{r}_i}{dt} = \vec{F}_i = \sum_j \vec{F}_{ij}, \tag{3}$$

which corresponds to particle motion in a very viscose liquid (this let us use more simple differential equations of first order instead of second order equations for standard Newtonian dynamics). To preserve the initial symmetry we take all viscosities μ_i=1.

With only attractive forces present our n point system eventually collapses towards the origin.
A collapse of all n points happening before the vertices belonging to "clusters in the network"' have separately concentrated in different regions defeats our goal of identifying the latter clusters.
To avoid the radial collapse we constrain $\vec{r}_i$, to be at all times on the unit sphere.
While the above avoids the radial collapse, the residual tangential forces can still initiate a collapse at some point on the unit sphere. After a sufficient time (or sufficiently many steps in evaluation of our dynamic system) has elapsed so that any point moved on average an appreciable distance away from its initial location geometrical clusters of points tend to form. The points in each geometrical cluster correspond to the original vertices in a cluster of the network which these points represent. (We recall the definition of a cluster in the graph/network as a subset nodes with a higher number of connections between them than the average number of connections with ``external" nodes, which are not in the cluster.)

Then, calculating the mutual distances between nodes, one can separate them into groups of the clusters. Therefore, scaling the parameter of the "critical" distance one can re-define clusters based on the intensity of connections, and to resolve sub-clusters of the clusters. It should be noted that definition of the function $g(C_{ij})$ gives the principal for a cluster definition: intensity of connections, e-mail exchange, etc.

For monitoring rapid changes of the network topology we calculate mutual entropies of the network (see for details refs.[10,12]). To do this we redefine the connectivity matrix in terms of probabilities of the connectivity in such way that each matrix element represents the probability that two nodes are connected to each other. Thus we normalize the connectivity matrix C so that

$$\sum_{i,j=1}^{n} C_{ij} = 1. \tag{4}$$

Then the sum over all columns $P_i = \sum_j^n C_{ij}$ can be considered as the probability of the connectivity for the node i, and the Shannon entropy

$$H(row) = -\sum_{j=1}^{n} P_i \log P_i \tag{5}$$

is a measure of the uncertainty of the connections for a given network. In the same way one can define the entropy for "inversed" connections: $H(column)$.(Due to symmetry of the connectivity matrix in our case, $H(row) = H(column)$.) The amount of mutual information (or negative entropy) gained via the given connectivity of the network is

$$\begin{aligned} I(C) &= H(row) + H(column) - H(column \mid row) \\ &= \sum_{i,j}^{n} C_{ij} \log(C_{ij} / P_i P_j), \end{aligned} \tag{6}$$

where

$$H(column \mid row) = -\sum_{i,j}^{n} C_{ij} \log(C_{ij}). \tag{7}$$

It should be noted that $I(C)$ does not depend on the vertex relabeling, and, as a consequence, this is a permutation invariant measure for the connectivity matrix. If the mutual entropies for two connectivity matrices are different, they represent different topological structures of network. This entropy is already sufficient to distinguish even between graphs that are normally cospectral.

The obvious extension of this definition of mutual Shannon entropy (information) could be used for calculations of mutual Rényi entropy . For example, based on definition of Rényi entropy [13], the expressions in eqs.(5) and (6) are
transformed into

$$H_q(row) = -\frac{1}{1-q} \log \sum_{j=1}^{n} P_i^q \tag{8}$$

and

$$H_q(column \mid row) = -\frac{1}{1-q} \sum_{i,j}^{n} \log(C^q{}_{ij}), \tag{9}$$

giving mutual information Rényi $I_q(C)$ for the given matrix C .

Using the Rényi information, one can not only distinguish between different network topologies on the base of the connectivity matrixes but extract information about network topology, such as number of clusters, cluster's dimensionalities etc. Moreover, by monitoring appropriate functions of mutual information, one can observe in real time a change in topology of the given network including a cluster formation, disappearance or appearance of group connections, change of the connection "styles", and other features.
For example, the difference between mutual Shannon information and Rényi information of kind 2 ($q = 2$) displays a sharp dependence of the size of the formed cluster (Figure 1).

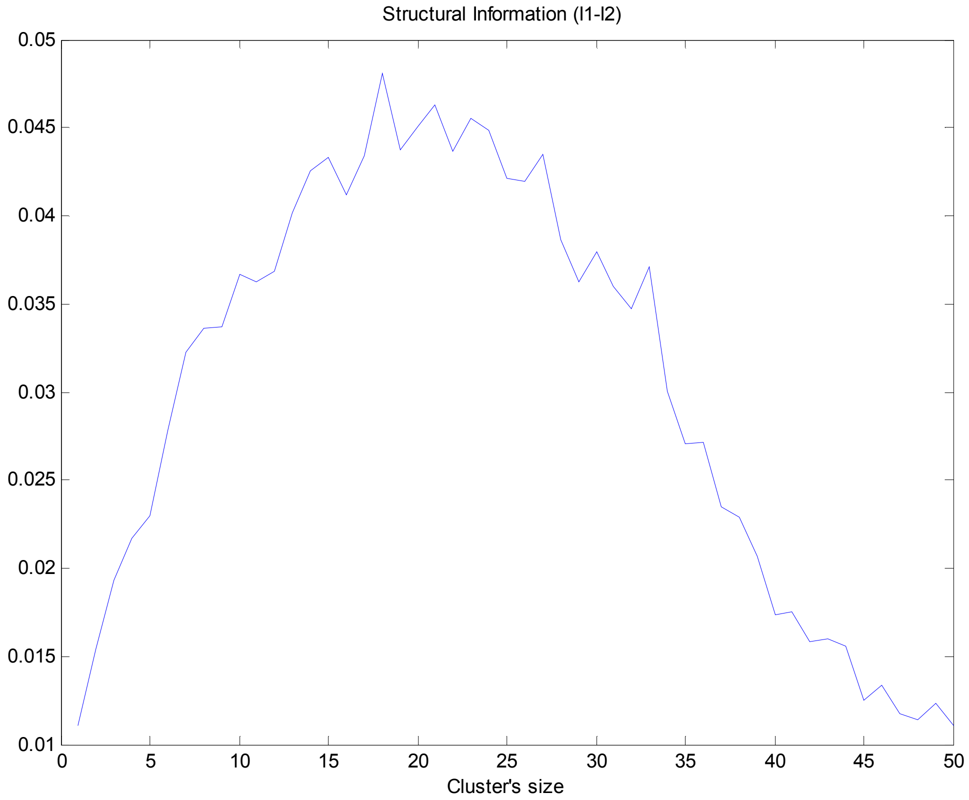


Figure 1: The difference between mutual Shannon information and Rényi information of kind 2.

This gives a unique opportunity to monitor dynamical behavior of the network in real time. It should be noted that different entropies are sensitive to different patterns of network topology (such a size of clusters, number of clusters, fractional dimensionality, etc), therefore many important properties of network can be extracted using suggested methods.

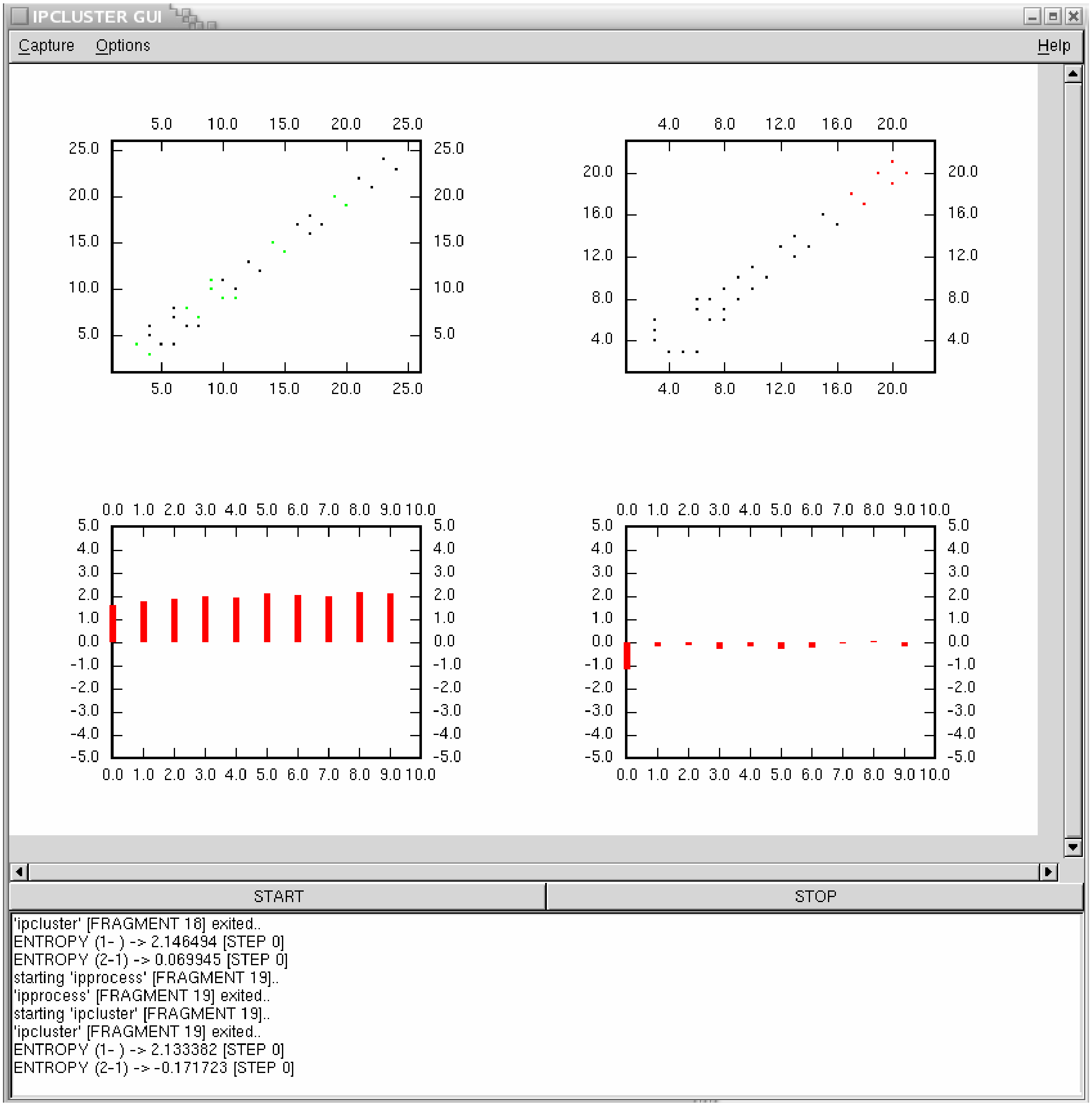
3.0 PROGRAM DESCRIPTION

Based on the described algorithms we have developed the real time network system “Ipcluster” for dynamical reconstruction and monitoring network communications. The program contains three modules: capture and connectivity matrix construction, analysis, and visualization.

The capture module is a program which puts the network card of the computer in promiscuous mode, captures the network packets and dumps the packets along with the specified headers parameters to a file (as a back-up). Then builds a connectivity matrix for all currently active nodes based on their IP addresses.

The analysis module applies the topology reconstruction (cluster identification) algorithm and calculates a set of Rényi entropies for the obtained connectivity matrix. In the current program the complexity of algorithms are and for the topology reconstruction and entropy calculations, correspondingly. Therefore, we can separate the time intervals for topology analysis and entropy monitoring, since the last one can be done in much less time.

The visualization module (see the typical snap short on Figure 2) presents both results of topology reconstruction and entropy on separate windows.



```
'ipcluster' [FRAGMENT 18] exited..
ENTROPY (1- ) -> 2.146494 [STEP 0]
ENTROPY (2-1) -> 0.069945 [STEP 0]
starting 'ipprocess' [FRAGMENT 19]..
'ipprocess' [FRAGMENT 19] exited..
starting 'ipcluster' [FRAGMENT 19]..
'ipcluster' [FRAGMENT 19] exited..
ENTROPY (1- ) -> 2.133382 [STEP 0]
ENTROPY (2-1) -> -0.171723 [STEP 0]
```

Figure 2: Visualization Module Snapshot

The two upper windows represent the reconstructed cluster structure of the network within the given time delay Δt. The cluster representation in each window uses different colors for continuously active connections during current Δt period; for connections which existed for the previous time interval, but disappeared for the current Δt; and for connections that just appeared for the current Δt period. Therefore, one can monitor topological changes on the network with the interval Δt.

Two lower windows present the histograms for chosen combinations of generalized mutual entropies as functions of time. It gives the opportunity to monitor particular topological structures in the network for which the given function of entropies is sensitive most. The upper windows are refreshed with the interval Δt, but the lower ones plot continuous histograms over time.

It should be noted that the visualization module dynamically analyze the reconstructed cluster structure for each Δt interval. By the end of each interval, it applies the reconstruction algorithm on the captured packets and displays the reconstructed topological structure. During the process of reconstruction, another thread to capture traffic packets for the next step is running. Therefore we do not lose any network traffic.

The current version of the program lets to choose a variety of options for definition of different actions as a connection, as frequency of connection, port of connection, protocol etc. Also we can vary time of monitoring and frequency of analysis, as well as different sets of mutual information to be visualized.

4.0 CONCLUSIONS

The presented program for real time network topology monitoring provides extremely fast method and confirms readability and efficiency of the approach developed in papers [9-13]. It may be used for real time monitoring of large networks and the internet. The presented algorithms may be also applied for dynamical monitoring and analysis of different kinds networks, for example communication networks, social networks etc. It provides quantitative method to define connected groups (clusters) one large networks with the ability to extract topology (structure and sub-structure of clusters) of the given network in real time with elements of visualization

BIBLIOGRAPHY

1. A. Reka, J. Hawoong and B. Albert-Laszlo, “Error and Attack Tolerance of Complex Networks”, Nature, Vol. 406, pp. 378-381, 2000.
2. S. H . Strogatz, “Exploring Complex Networks”, Nature, Vol. 410, pp. 268-276, 2000.
3. L. Deri L. and S. Suin, “Practical Network Security: Experiences with ntop”, Computer Networks, Vol.34, pp. 873-880, 2000.
4. P. A. Porras and A. Valdes, ``Live Traffic Analysis of TCP/IP Gateways", Internet, Society Symposium on Network and Distributed System Security, SanDiego, California March 11-13, 1998.
5. J. B. D. Cabrera, B. Ravichandram and R. K. Mehra, “Statistical Traffic Modeling for Network Intrusion Detection", Proceedings of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, IEEE, 2000.
6. T. Huisinga, R. Barlovic, W. Knospe, A. Schadschneider, and M. Schreckenberg, “A Microscopic Model for Packet Transport in the Internet”, arXiv:cond-mat/0102516, 2001.

7. N. G. Duffield and Grossglauser,"Trajectory Sampling for Direct Traffic Obsevation", MIEEE/ACM Transactions on Networking, Vol. 9, No 3, pp. 280-292, 2001.
8. M. Ayedemir, L. Bottomley, M. Coffin, C. Jeffries, P. Kiessler, K. Kumar, W. Ligon, J. Marin, A. Nilsson, J. McGovern, A. Rindos, K. Vu, S. Woolet, A. Zaglow, K. Zhu, "Two Tools for Network Traffic Analysis", Computer Networks, Vol. 36, pp.169-179, 2001.
9. V. Gudkov, J. E. Johnson and S. Nussinov, "Approaches to Network Classification", arXiv: cond-mat/0209111 (2002); submitted to Phys. Rev. E.
10. V. Gudkov and S. Nussinov, "Graph equivalence and characterization via a continuous evolution of a physical analog ", arXiv: cond-mat/0209112 (2002).
11. V. Gudkov, S. Nussinov and Z. Nussinov, "A Novel Approach Applied to the Largest Clique Problem", arXiv: cond-mat/0209419 (2002).
12. V. Gudkov and J. E. Johnson, "Applications of Generalized Entropies to Network Analysis", talk at the "Networks 2003", Santa Fe, NM (2003).
13. A. Rényi, "Probability Theory", North-Holland Pub. Co. – Amsterdam, London & American Elsevier Pub. Co., Inc.-New York (1970).

Monitoring of Network Topology Dynamics

Information Systems Technology Panel Symposium
on
Adaptive Defense in Unclassified Networks
Toulouse France
April 19-20, 2004

Vladimir Gudkov PhD, Joseph E. Johnson PhD
Rajesh Madamanchi, James L. Sidoran

Authors
(presented by Joseph E. Johnson PhD)

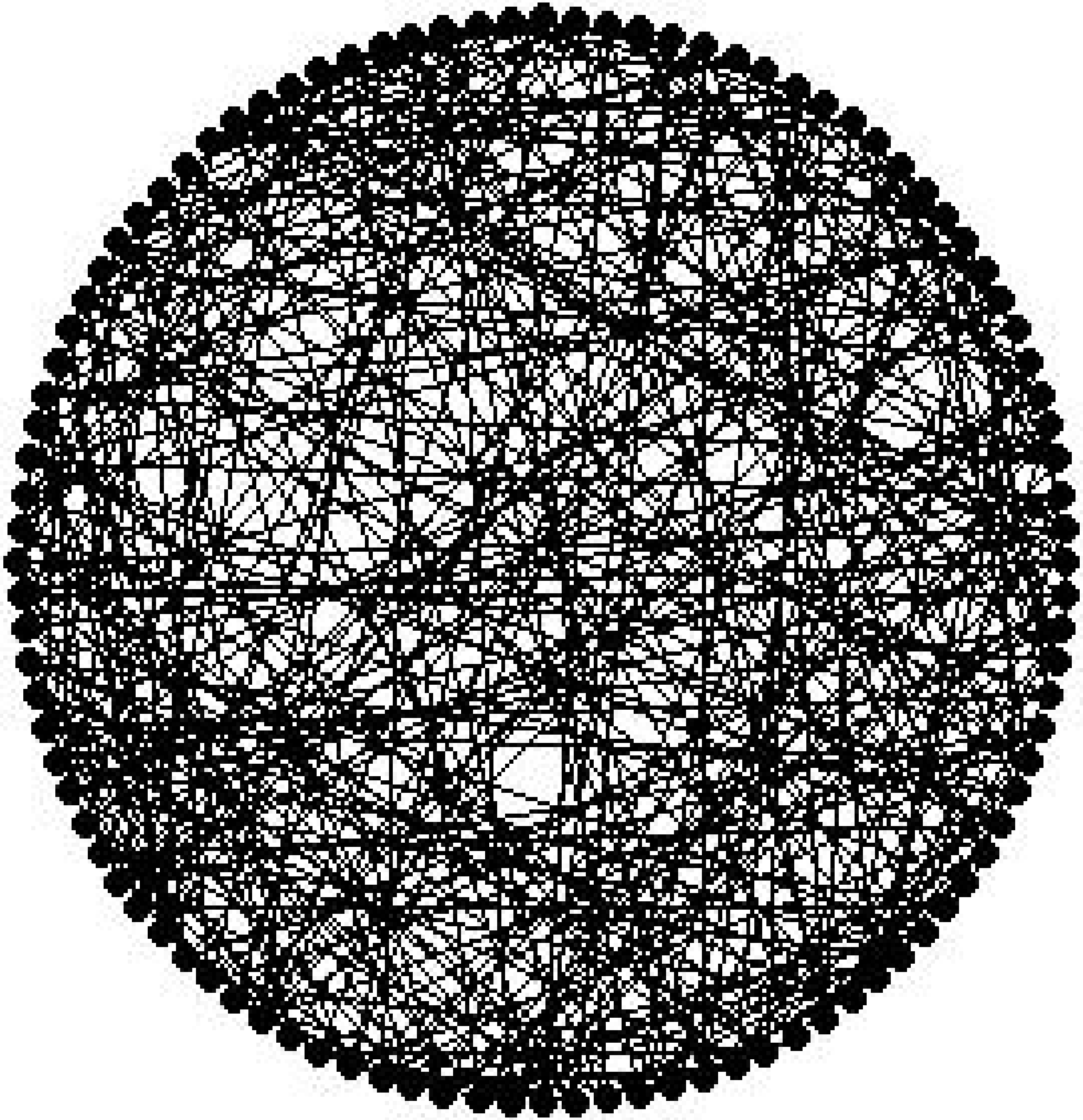
- **Dr. Vladimir Gudkov / Dr. Joseph Johnson**
  - **Department of Physics and Astronomy**
  - **University of South Carolina,**
  - **Columbia, SC 29208**
  - **USA**
  - **Email: gudkov@sc.edu / jjohnson@sc.edu**
- **Mr. Rajesh Madamanchi**
  - Department of Computer Science and Engineering
  - University of South Carolina,
  - Columbia, SC 29208
  - USA
  - Email: madamanc@engr.sc.edu
- **Mr. James L. Sidoran**
  - Air Force Research Laboratory
  - Defensive Information Warfare
  - Rome, NY 13441-4505
  - USA
  - Email: James.Sidoran@rl.af.mil

Network Topology – Introduction 1

- Pervasive nature of networks
  - Communication (phone & internet)
  - Transportation (air, highway, rail, water, pipeline)
  - Utility (water, sewer, power)
  - Distribution (supply and distribution)
  - Social (organizational, relationships)
  - Biological (neural, fluid)

Network Topology – Introduction 2

- Networks are exceptionally complex & difficult
  - The combinatorics becomes overwhelming
    - For any topology there are n! ways to relabel the nodes
  - Identical topologies not easily discerned
    - n! relabeled connectivity matrices must be compared
  - Classifications are usually isospectral
    - (eignenvalues of different topologies are the same)
  - Difficulties reside in the symmetry under the permutation (symmetric) group



Network Topology - Introduction 3

- One cannot see the forest for the trees
  - A network of 1 Million nodes is represented by a connectivity matrix with 1 trillion ones and zeros or a trillion lines in the graph shown – all of apparently the same value
  - A hierarchical analysis or classification is not available to classify networks dominant to less dominant.
- There is nothing similar to satellite photography of the ground viewed at different resolutions

Network Topology - Introduction 4

- Internet
  - Extremely large number of nodes
  - Critically important to optimize and maintain
  - Extremely important to protect systems from intrusions, attacks, and aberrant behavior
  - Yet there is not yet an adequate mathematical foundation to solve these problems
  - Currently we solve these problems more from experience and practical techniques.

Our Approach - 1

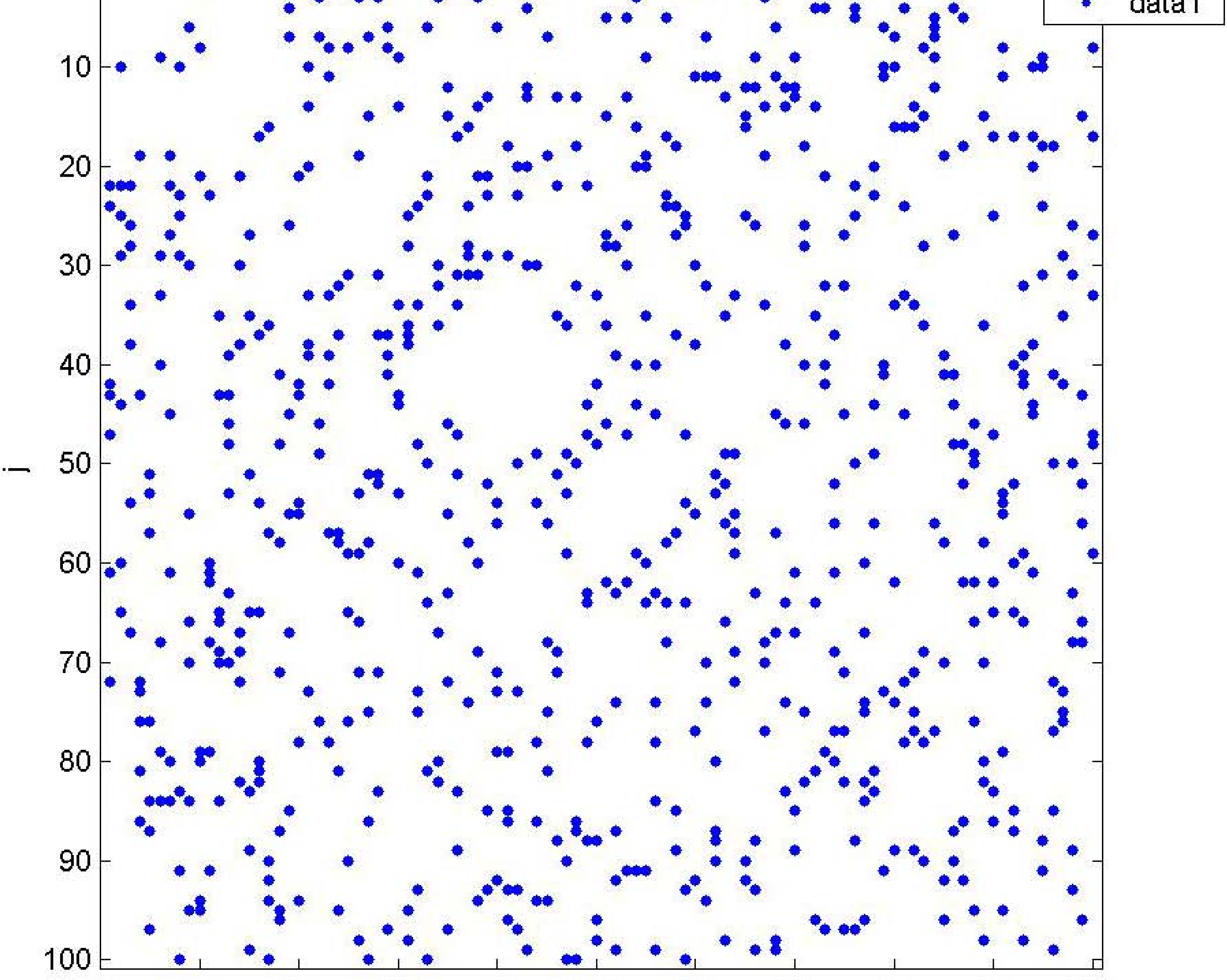
- Our approach: If we could identify the dominant changes in the topology we might could better optimize, manage, and thwart intrusions.
- Dominant changes implies we have a hierarchical approach of more to less important)
- Then aberrant topologies may be linked with undesirable behavior or attacks

Our Approach - 2

- Our techniques to understand topology
  - Clusters – how to identify
  - Entropy & generalized Renyi entropy
  - Entropy differences
  - Fast algorithms to track large systems in time
- Our more recent techniques
  - Lie Group Theory & Markov Processes

Connectivity Matrix

- One can exactly specify a network by the connectivity or adjacency matrix defined by C_{ij} = 1 if nodes i & j are connected, and = 0 otherwise.
- It is important to note that the diagonal can be set to 0 (if a thing is considered not connected to itself) or 1 (if it is).
- Thus the diagonal is arbitrary


data1
j
10
20
30
40
50
60
70
80
90
100

A Cluster Identification Algorithm

- Use symmetrical initial configuration of network as equally spaced nodes in n-1 dimensions (hyper-tetrahedron)
- Connected nodes have a ‘force of attraction while unconnected nodes have a force of repulsion) with 1 order dynamics
- Iterate the system to ‘condense’ clusters at hierarchical levels.

Dynamical Tracking

- Use clusters to identify groups of nodes for the computation of generalized entropies of the form: a*log(b(P^n))
- Track the entropy of different clusters over time along with the entropy differences
- Identify large topological changes with aberrant changes in the entropy of component clusters.

Capture Program

- We have built a lightweight version of a C_{ij} capture program (snort type program) which can be easily adjusted to capture the routing information from servers and build the associated C_{ij} rapidly.
- Display techniques were developed to monitor these changes over time.

Primary Results

- Although all of the different entropies are sensitive to cluster changes over time, the entropy differences seem to be the most sensitive.
- Our current work is in the collection of data on correlating aberrant behavior with generalized entropies of identified clusters and entropy differences.

DARPA Funded Initiative

- University of South Carolina
  - Complex Problems Group
  - 5 past DARPA Awards
  - A new award is pending

New Work Uses Markov Type Lie Groups

Techniques for the problems

- We are using methods from
  - Lie Groups & Lie Algebras
  - Markov & General Linear Transformations
  - New interpretation of eigenvalues
  - Non-recurring paths identified
  - Hierarchical expansion of an ‘information function’

‘System’ space & dynamics

How do we describe the network system space – Compare to physics.

  - What is the ‘space’ (like 3 dim. for particles)
  - What are the metrics (momentum, force..)
  - What are the dynamical laws.

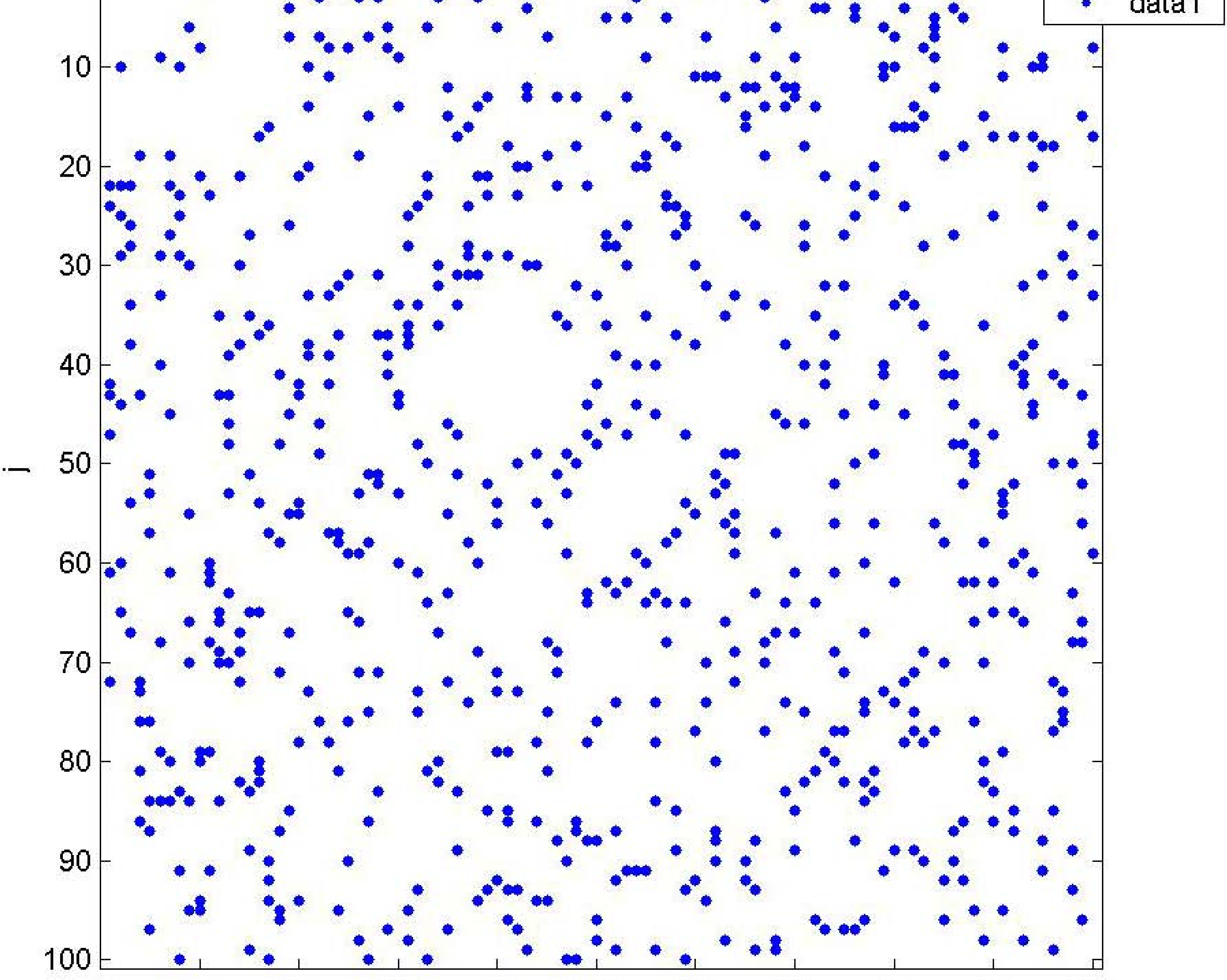
- The state of the system is a matrix C_{ij} and not a vector

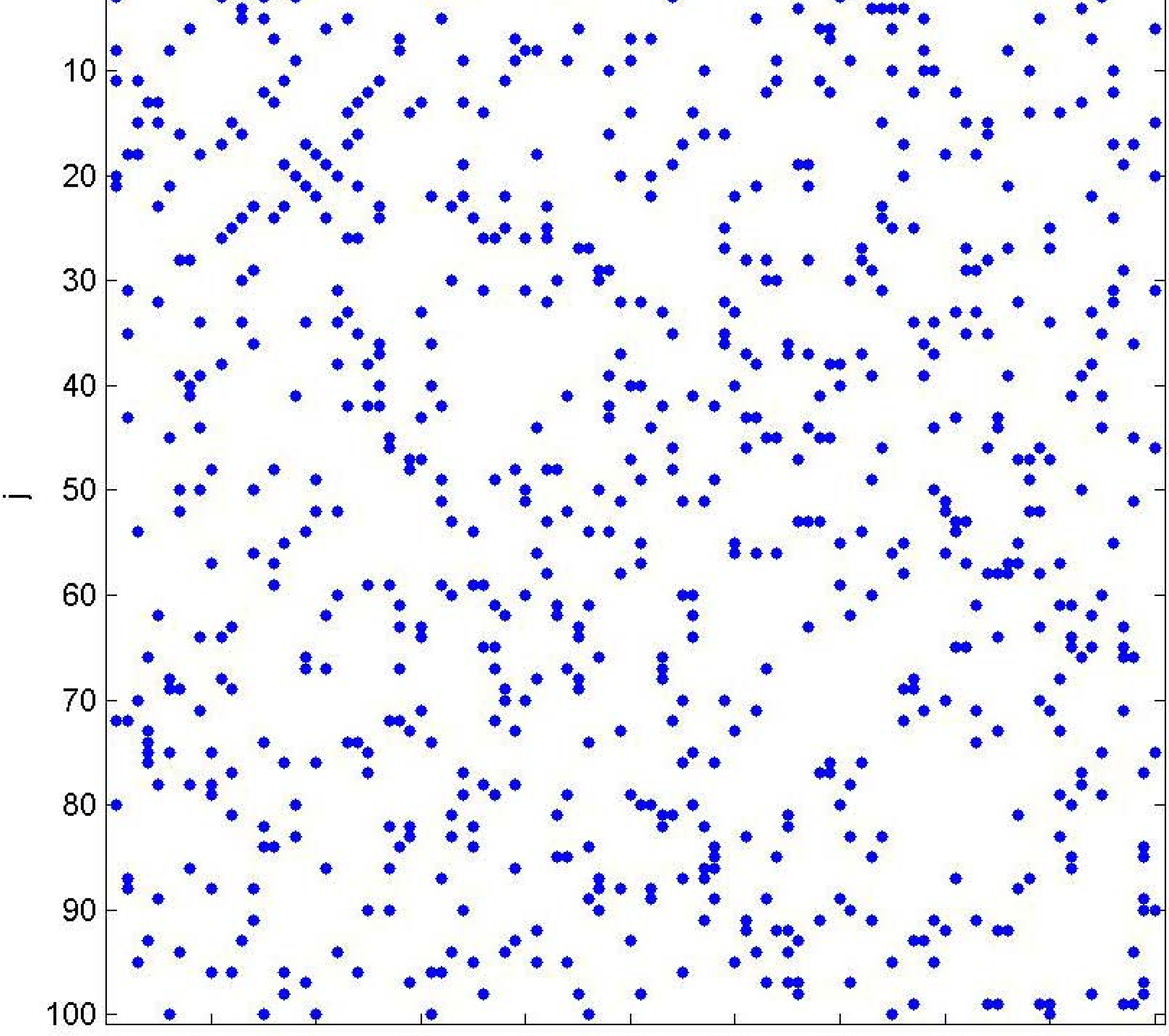
Linkage of Markov Groups & Networks

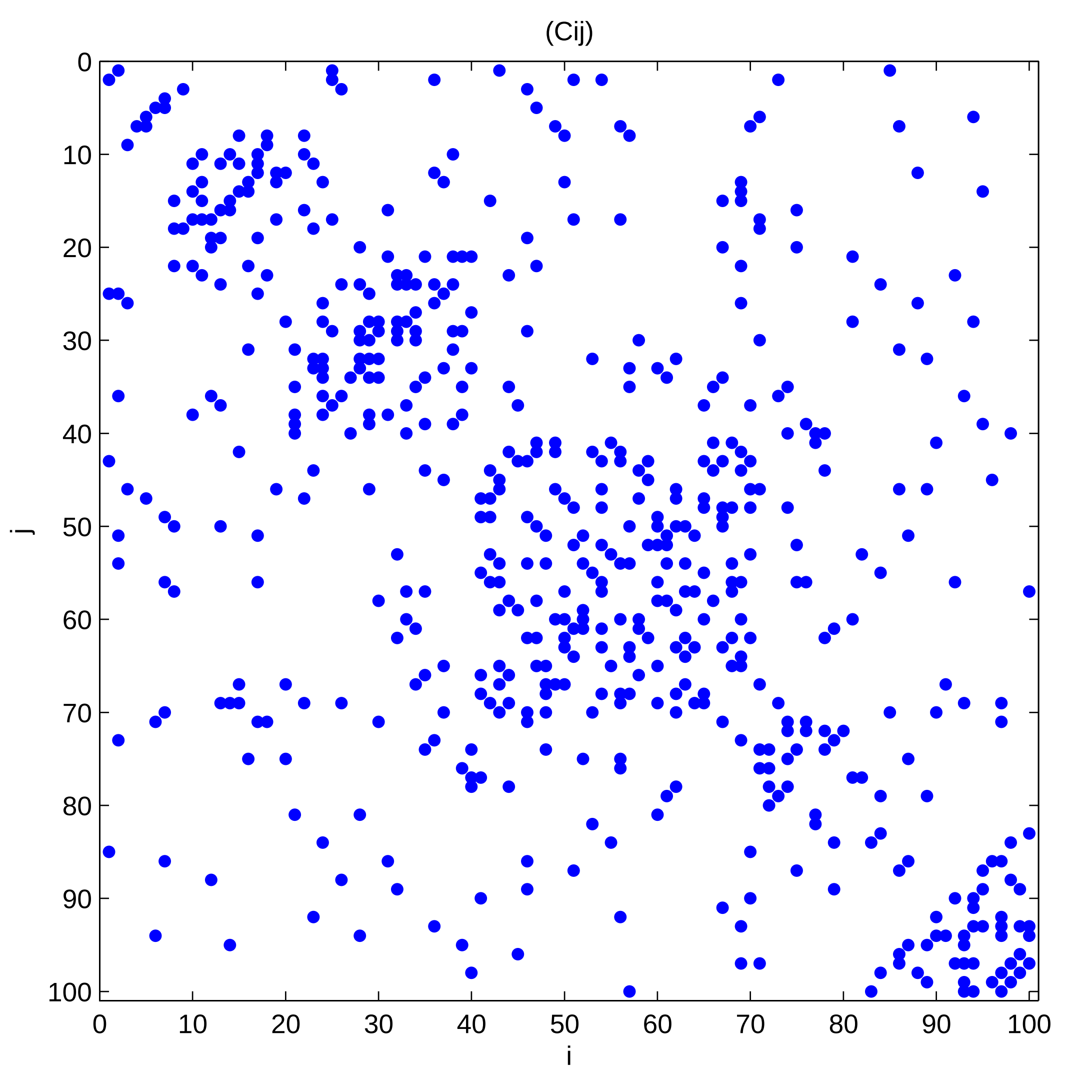
- Observation: C_{ij} is an element of the Markov Lie Algebra (monoid) (when the diagonal = negative sum of column).
  - C_{ij} generates conserved flows of a conserved quantity or entity (information, water, goods, people ….).
  - Other diagonals represent GL(n,R) transformations with birth/decay of quantity at the nodes at fixed rates.
  - This allows the power of Lie groups and Lie algebras
  - A static problem is now seen as dynamic
  - Multiple areas of mathematics are now connected

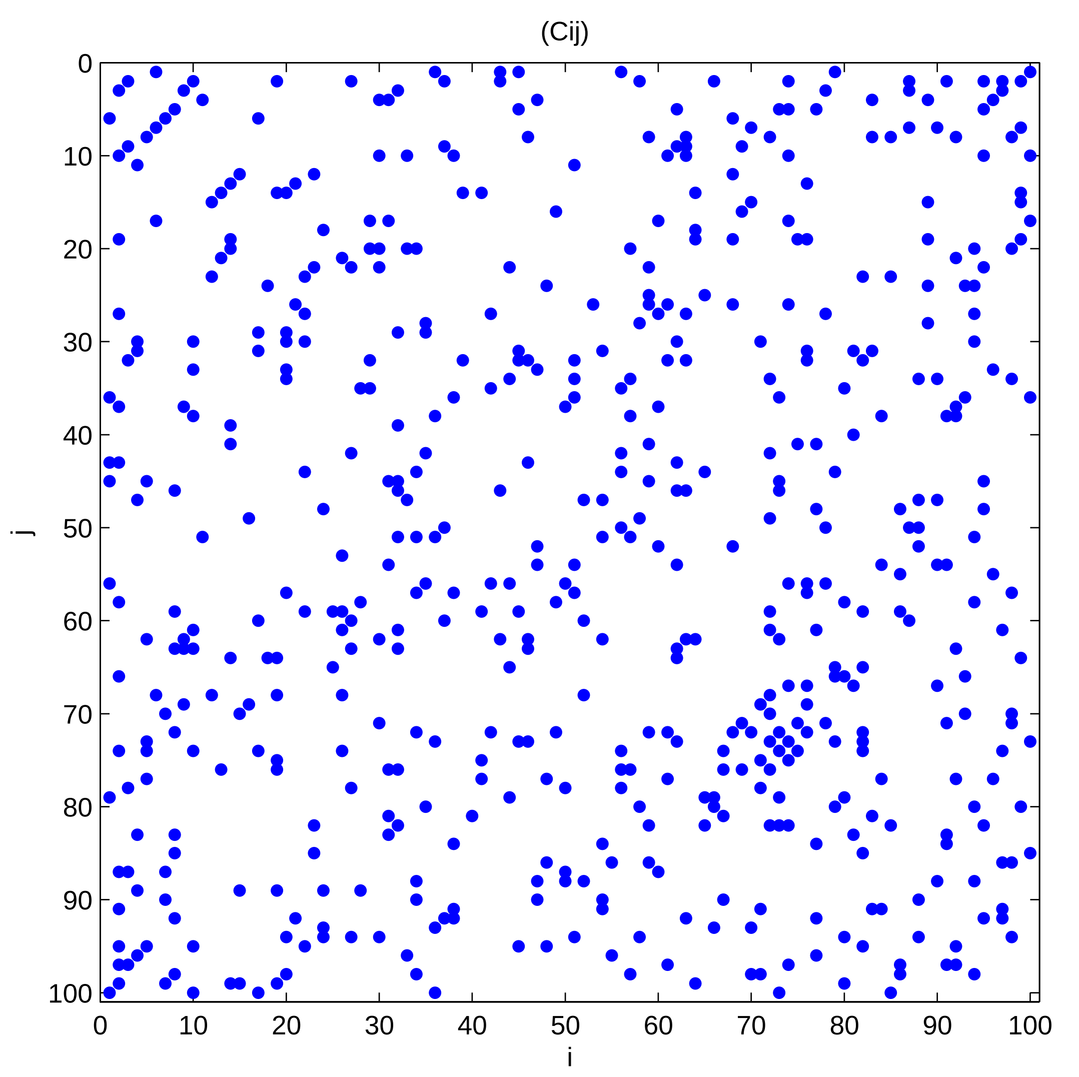
C_{ij} meaning

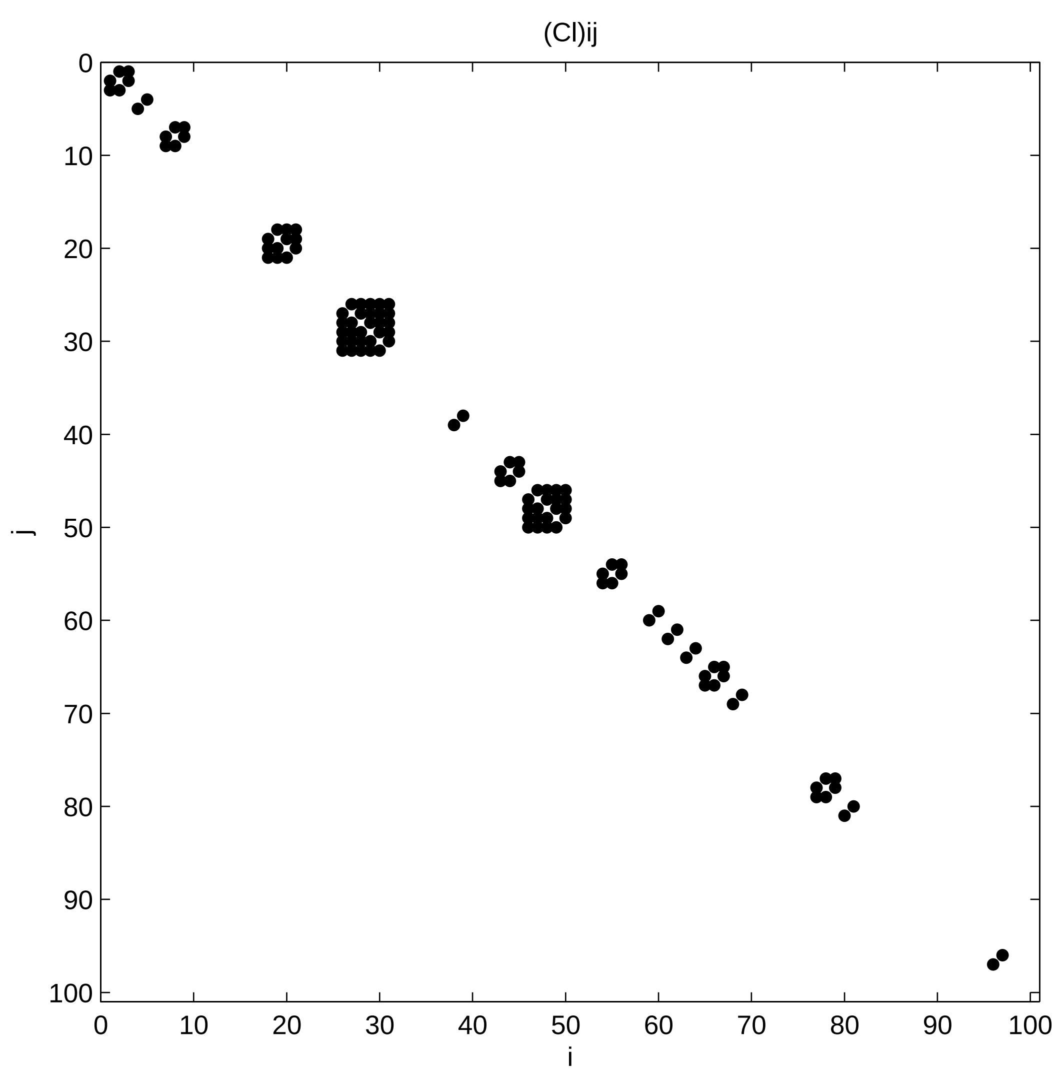
- Does this help give meaning to the eigenvectors & eigenvalues of C_{ij} ?
- Yes: This group theory approach gives meaning to the eigenvalues as exponential rates of approach to equilibrium
- The eigenvectors are those combinations of nodes with a unique information flow rate (the associated eigenvalue) in an approach to equilibrium.

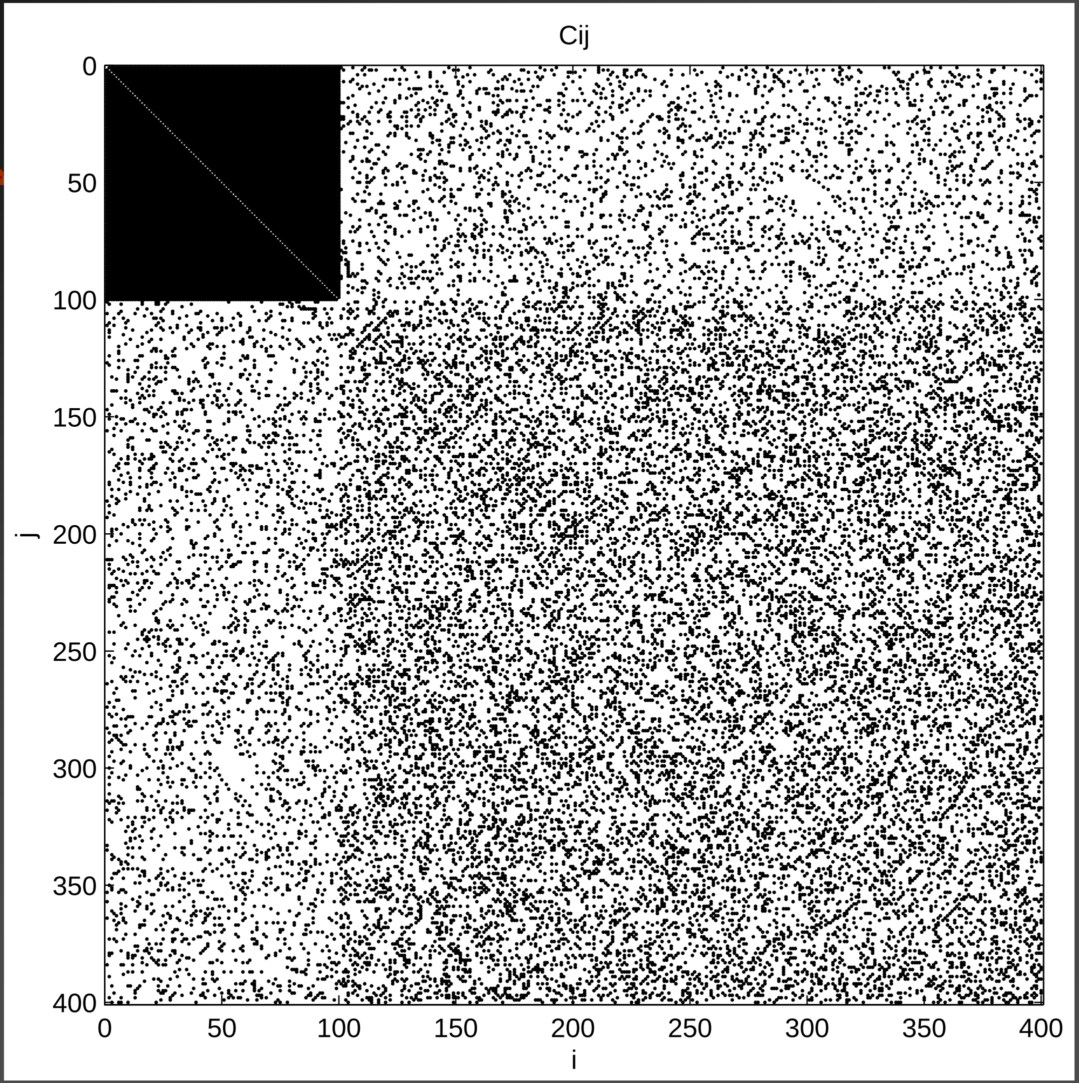

data1
j
10
20
30
40
50
60
70
80
90
100

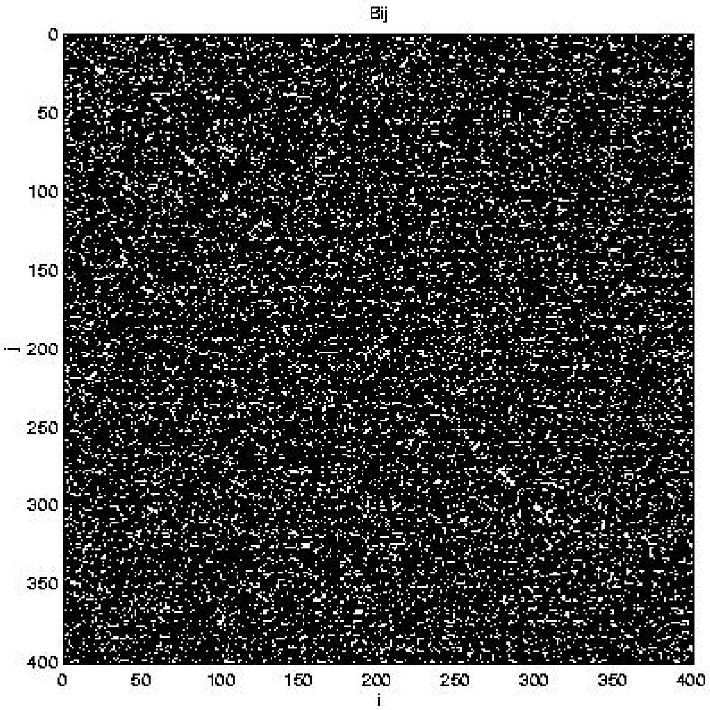

j
10
20
30
40
50
60
70
80
90
100

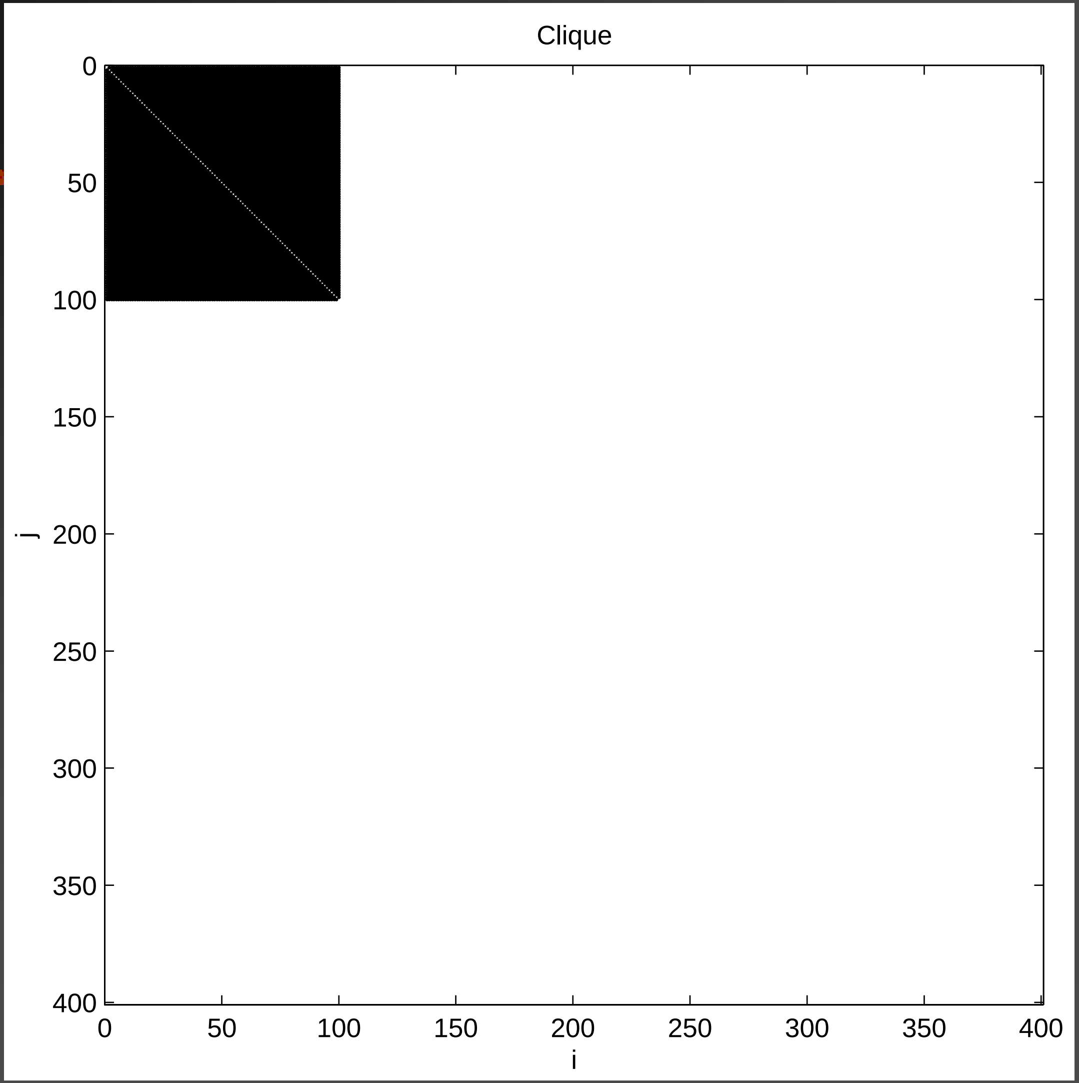

(Cij)
j
i
0
10
20
30
40
50
60
70
80
90
100

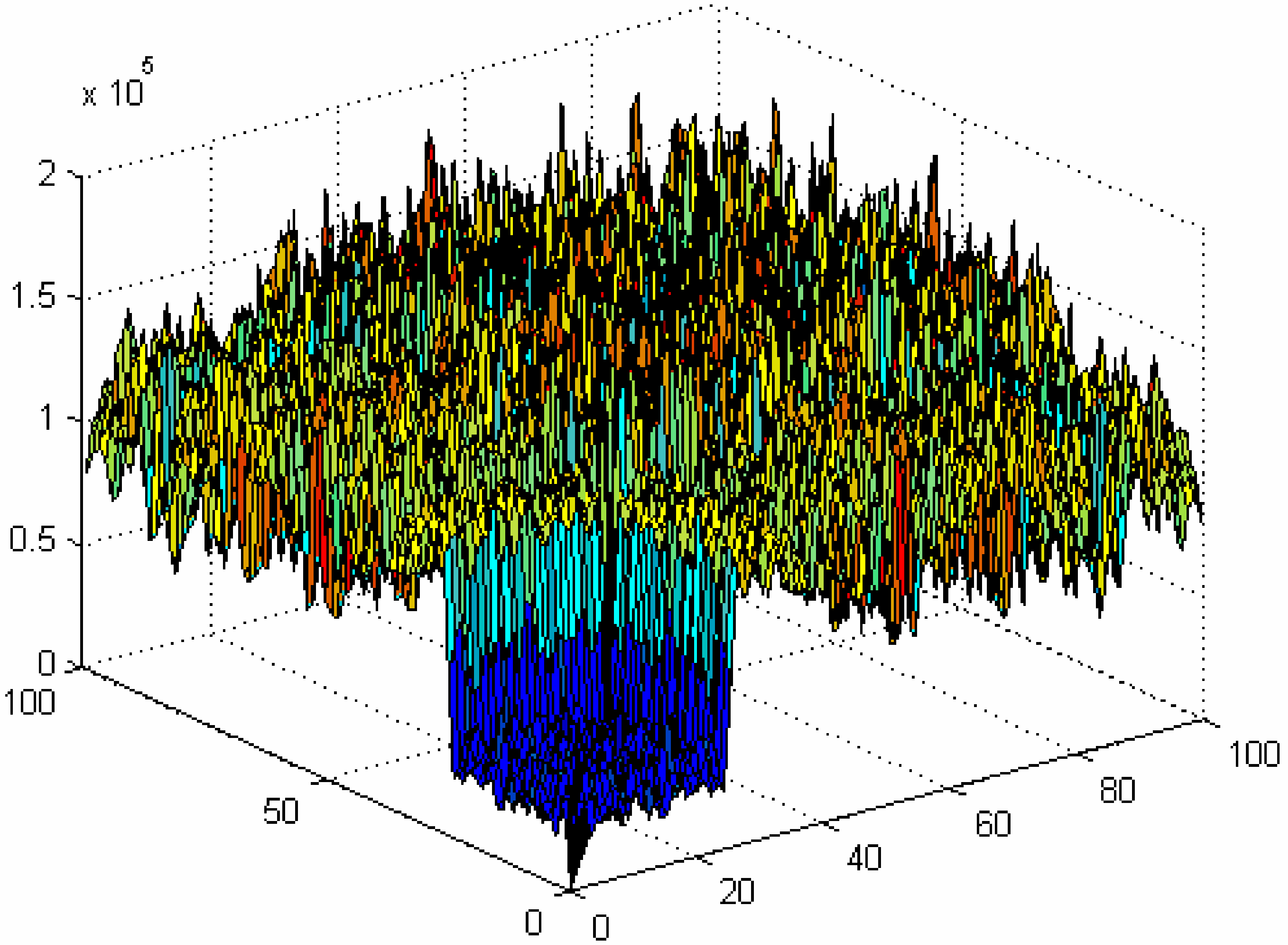

(Cij)
0
10
20
30
40
50
60
70
80
90
100
j
0
10
20
30
40
50
60
70
80
90
100
i

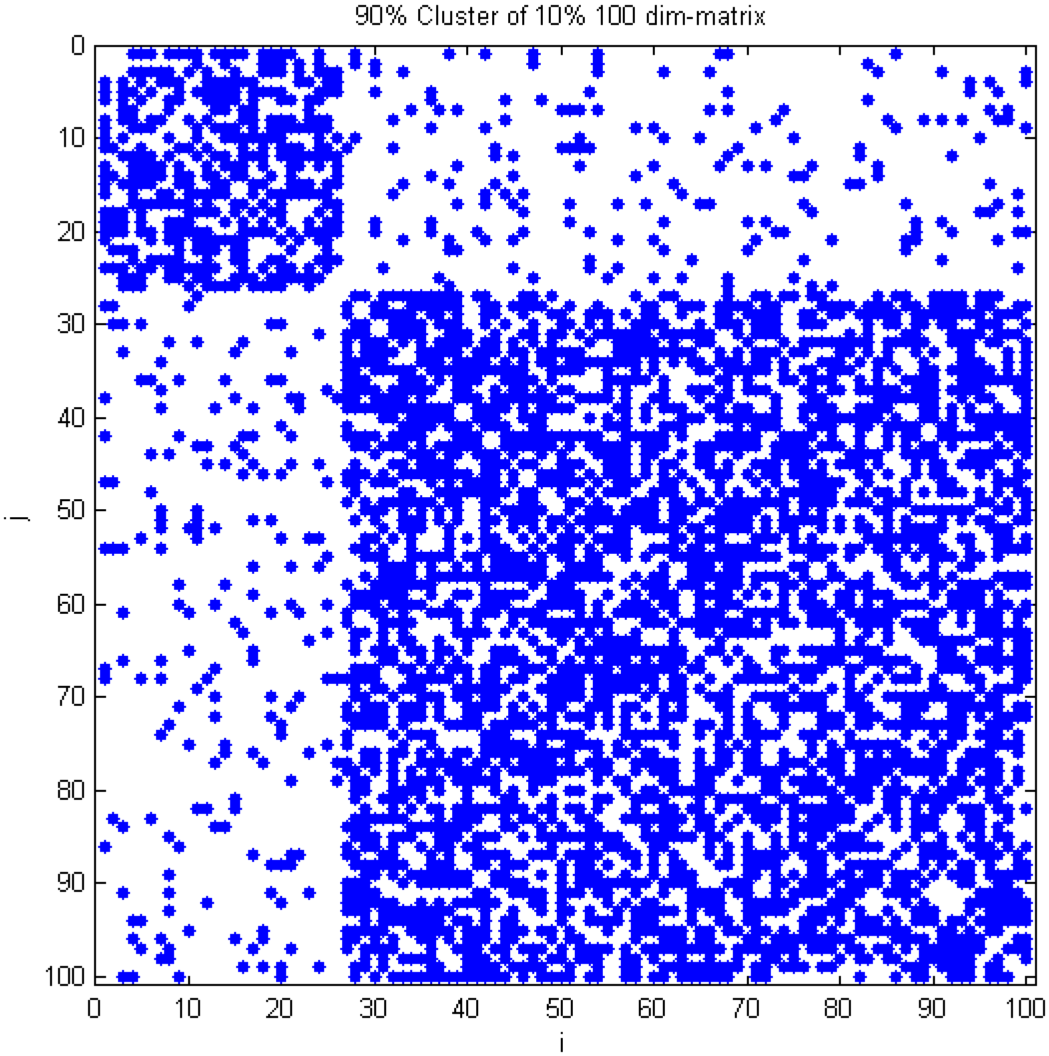

(Cl)ij
0
10
20
30
40
50
60
70
80
90
100
j
0
10
20
30
40
50
60
70
80
90
100
i

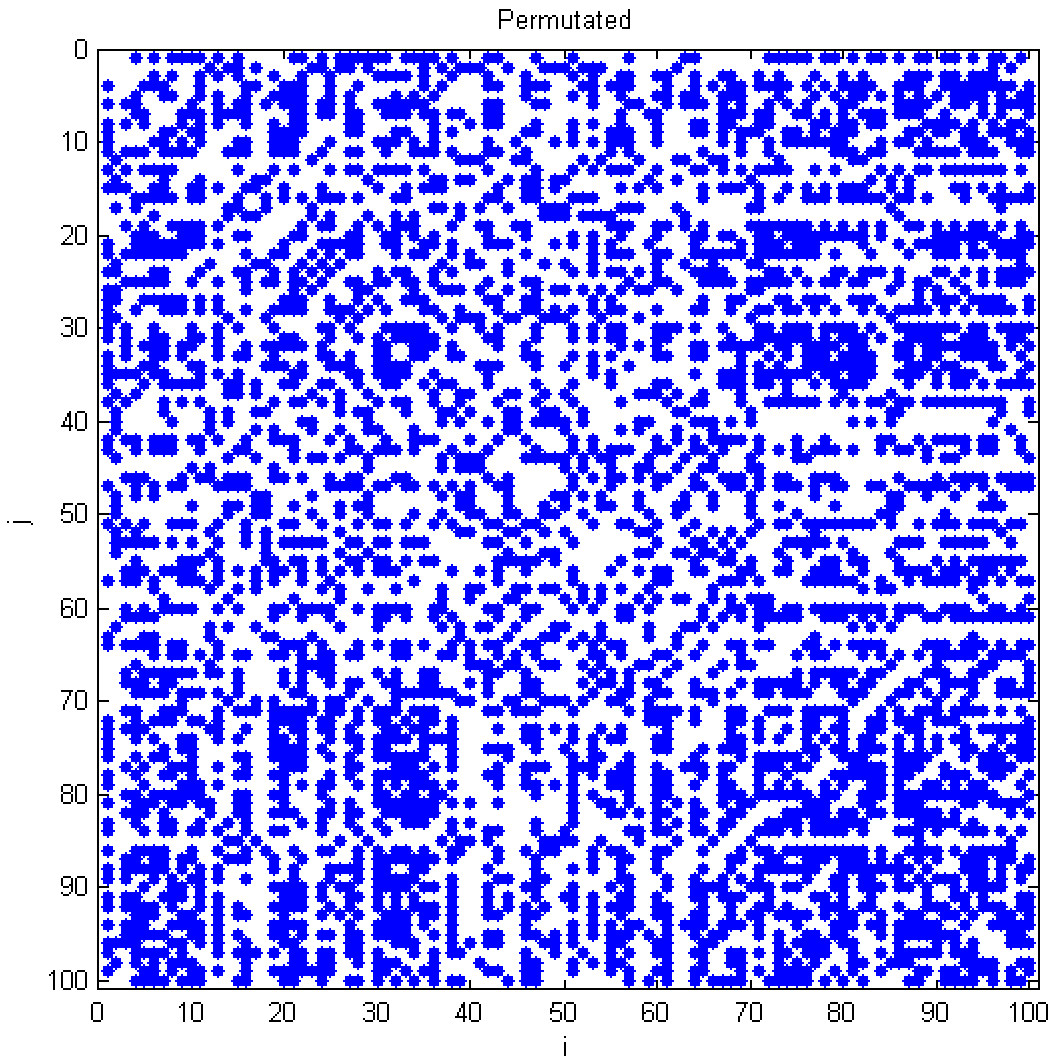

Cij
0
50
100
150
200
250
300
350
400
j
0
50
100
150
200
250
300
350
400
i

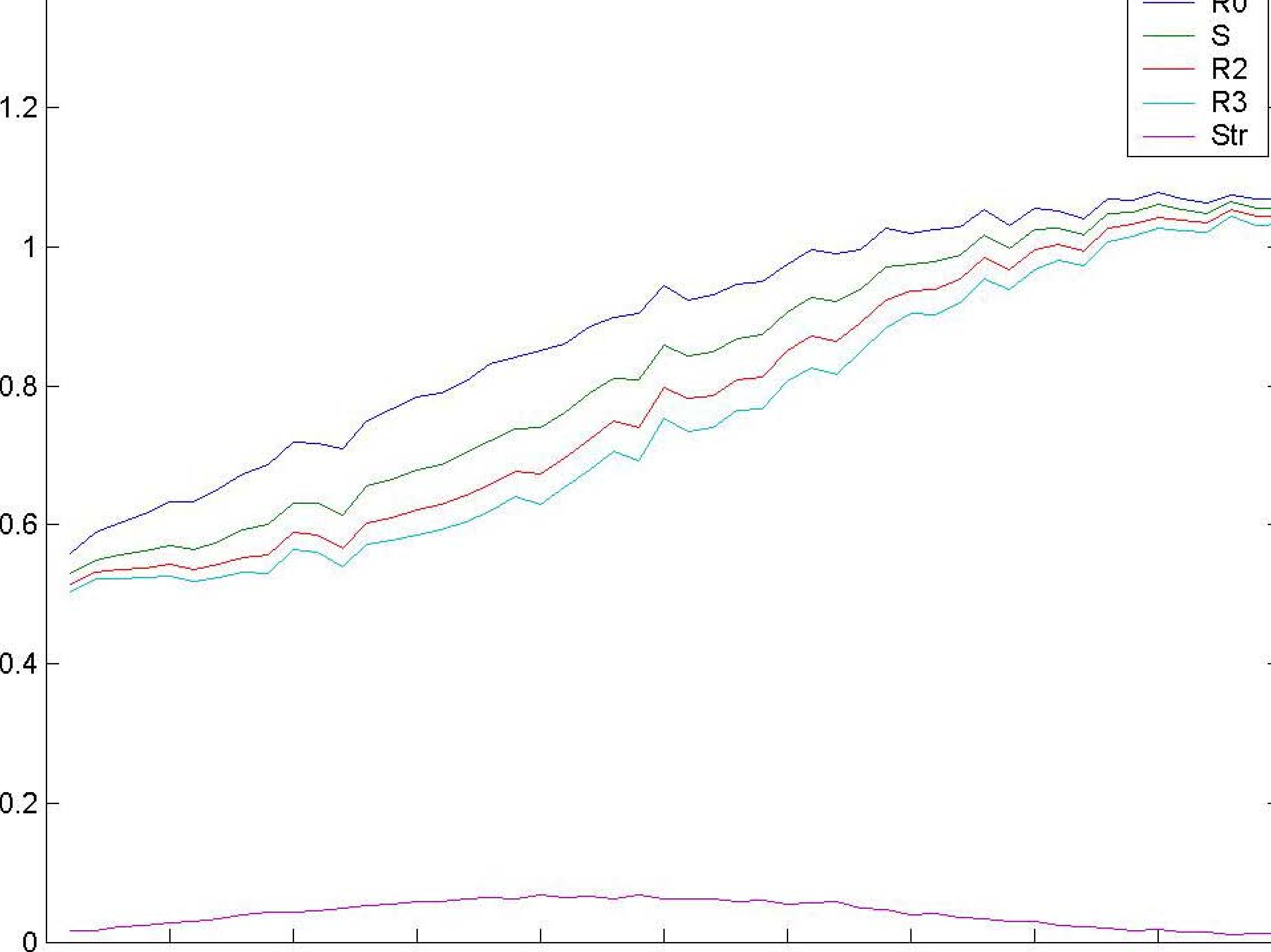

Bij
0
50
100
150
200
250
300
350
400
j
0
50
100
150
200
250
300
350
400
i

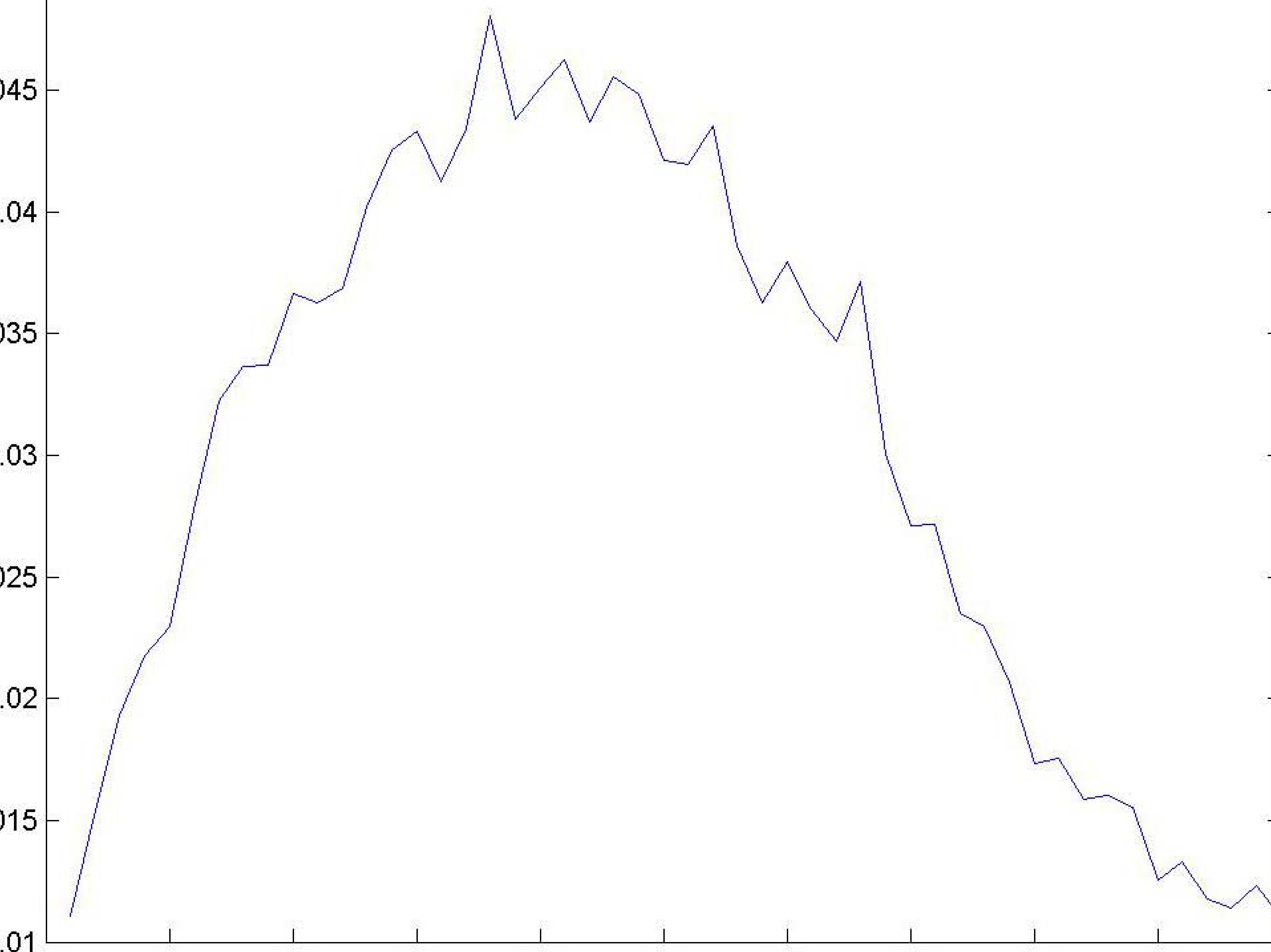

Clique
0
50
100
150
200
250
300
350
400
j
0
50
100
150
200
250
300
350
400
i


x 10^5
2
1.5
1
0.5
0
100
50
0
0
20
40
60
80
100


90% Cluster of 10% 100 dim-matrix
j
0
10
20
30
40
50
60
70
80
90
100
i
0
10
20
30
40
50
60
70
80
90
100


Permutated
j
0
10
20
30
40
50
60
70
80
90
100
i
0
10
20
30
40
50
60
70
80
90
100


R0
S
R2
R3
Str
1.2
1
0.8
0.6
0.4
0.2
0


045
.04
035
.03
025
.02
015
.01

Thank You

- jjohnson@sc.edu